

Bernoulli 18(2), 2012, 678–702

DOI: [10.3150/10-BEJ344](https://doi.org/10.3150/10-BEJ344)

Estimation in semi-parametric regression with non-stationary regressors

JIA CHEN^{*}, JITI GAO^{**} and DEGUI LI[†]

School of Economics, The University of Adelaide, Adelaide SA 5005, Australia.

E-mail: ^{*}jia.chen@adelaide.edu.au; ^{**}jiti.gao@adelaide.edu.au; [†]degui.li@adelaide.edu.au

In this paper, we consider a partially linear model of the form $Y_t = X_t^\tau \theta_0 + g(V_t) + \epsilon_t$, $t = 1, \dots, n$, where $\{V_t\}$ is a β null recurrent Markov chain, $\{X_t\}$ is a sequence of either strictly stationary or non-stationary regressors and $\{\epsilon_t\}$ is a stationary sequence. We propose to estimate both θ_0 and $g(\cdot)$ by a semi-parametric least-squares (SLS) estimation method. Under certain conditions, we then show that the proposed SLS estimator of θ_0 is still asymptotically normal with the same rate as for the case of stationary time series. In addition, we also establish an asymptotic distribution for the nonparametric estimator of the function $g(\cdot)$. Some numerical examples are provided to show that our theory and estimation method work well in practice.

Keywords: asymptotic theory; nonparametric estimation; null recurrent time series; semi-parametric regression

1. Introduction

During the past two decades, there has been much interest in various nonparametric and semi-parametric techniques to model time series data with possible nonlinearity. Both estimation and specification testing problems have been systematically examined for the case where the observed time series satisfy a type of stationarity. For more details and recent developments, see Robinson [26–28], Fan and Gijbels [8], Härdle *et al.* [15, 16], Fan and Yao [9], Gao [10], Li and Racine [21] and the references therein.

As pointed out in the literature, the stationarity assumption seems too restrictive in practice. For example, when tackling economic and financial issues from a time perspective, we often deal with non-stationary components. In reality, neither prices nor exchange rates follow a stationary law over time. Thus practitioners might feel more comfortable avoiding restrictions like stationarity for processes involved in economic time series models. There is much literature on parametric linear and nonlinear models of non-stationary time series, but very little work has been done in nonparametric and semi-parametric nonlinear cases. In nonparametric estimation of nonlinear regression and autoregression of non-stationary time series models and continuous-time financial models, existing studies

This is an electronic reprint of the original article published by the ISI/BS in *Bernoulli*, 2012, Vol. 18, No. 2, 678–702. This reprint differs from the original in pagination and typographic detail.

include Phillips and Park [25], Karlsen and Tjøstheim [20], Bandi and Phillips [1], Karlsen *et al.* [19], Schienle [30] and Wang and Phillips [32, 33]. Recently, Gao *et al.* [11, 12] considered nonparametric specification testing in both autoregression and cointegration models.

Consider a nonparametric regression model of the form

$$Y_t = m(Z_t) + \epsilon_t, \quad t = 1, \dots, n, \quad (1.1)$$

where $\{Y_t\}$ and $\{Z_t\}$ are non-stationary time series, $m(\cdot)$ is an unknown function defined in \mathbf{R}^p and $\{\epsilon_t\}$ is a sequence of strictly stationary errors. We may apply a nonparametric method to estimate $m(\cdot)$,

$$\hat{m}(z) := \hat{m}_n(z) = \sum_{t=1}^n a_{nt}(z) Y_t, \quad (1.2)$$

where $\{a_{nt}(z)\}$ is a sequence of positive weight functions; see Karlsen *et al.* [19] and Wang and Phillips [32, 33].

As pointed out in the literature for the case where the dimension of $\{Z_t\}$ is larger than three, $m(\cdot)$ may not be estimated by $\hat{m}(z)$ with reasonable accuracy due to “the curse of dimensionality”. The curse of dimensionality problem has been clearly illustrated in several books, such as Silverman [31], Hastie and Tibshirani [17], Green and Silverman [13], Fan and Gijbels [8], Härdle *et al.* [15], Fan and Yao [9] and Gao [10]. There are several ways to circumvent the curse of dimensionality. Perhaps one of the most commonly used methods is semi-parametric modelling, which is taken to mean partially linear modelling in this context. In this paper, we propose using a partially linear model of the form

$$Y_t = X_t^\tau \theta_0 + g(V_t) + \epsilon_t, \quad t = 1, \dots, n, \quad (1.3)$$

where θ_0 is an unknown d -dimensional vector; $g(\cdot)$ is some continuous function; $\{X_t = (x_{t1}, \dots, x_{td})^\tau\}$ is a sequence of either stationary or non-stationary regressors, as assumed in A1 below; $\{V_t\}$ is a β null recurrent Markov process (see Section 2 below for detail); and $\{\epsilon_t\}$ is an error process. As discussed in Section 3.2 below, $\{\epsilon_t\}$ can be relaxed to be either stationary and heteroscedastic or non-stationary and heteroscedastic.

An advantage of the partially linear approach is that any existing information concerning possible linearity of some of the components can be taken into account in such models. Engle *et al.* [7] were among the first to study this kind of partially linear model. It has been studied extensively in both econometrics and statistics literature. With respect to development in the field of semi-parametric time series modelling, various estimation and testing issues have been discussed for the case where both $\{X_t\}$ and $\{V_t\}$ are strictly stationary (see, e.g., Härdle *et al.* [15] and Gao [10]) since the publication of Robinson [27]. For the case where $\{V_t\}$ is a sequence of either fixed designs or strictly stationary regressors but there is some type of unit root structure in $\{X_t\}$, existing studies, such as Juhl and Xiao [18], have discussed estimation and testing problems.

To the best of our knowledge, the case where either $\{V_t\}$ is a sequence of non-stationary regressors or both $\{X_t\}$ and $\{V_t\}$ are non-stationary has not been discussed in the literature. This paper considers the following two cases: (a) where $\{X_t\}$ is a sequence of strictly stationary regressors and $\{V_t\}$ is a sequence of non-stationary regressors; and

(b) where both $\{X_t\}$ and $\{V_t\}$ are non-stationary. In this case, model (1.3) extends some existing models (Robinson [27], Härdle *et al.* [15], Juhl and Xiao [18] and Gao [10]) from the case where $\{V_t\}$ is a sequence of strictly stationary regressors to the case where $\{V_t\}$ is a sequence of non-stationary regressors. Since the invariant distribution of the β null recurrent Markov process $\{V_t\}$ does not have any compact support, however, the semi-parametric technique used in stationary time series cannot be directly applicable to our case. In this paper, we will develop a new semi-parametric estimation method to address such new technicalities when establishing our asymptotic theory.

The main objective of this paper is to derive asymptotically consistent estimators for both θ_0 and $g(\cdot)$ involved in model (1.3). In a traditional stationary time series regression problem, some sort of stationary mixing condition is often imposed on the observations (X_t, V_t) to establish asymptotic theory. In this paper, it is interesting to find that the proposed semi-parametric least-squares (SLS) estimator of θ_0 is still asymptotically normal with the same rate as that in the case of stationary time series when certain smoothness conditions are satisfied. In addition, our nonparametric estimator of $g(\cdot)$ is also asymptotically consistent, although the rate of convergence, as expected, is slower than that for the stationary time series case.

The rest of the paper is organized as follows. The estimation method of θ_0 and $g(\cdot)$ and some necessary conditions are given in Section 2. The main results and some extensions are provided in Section 3. Section 4 provides a simulation study. An analysis of an economic data set from the United States is given in Section 5. An outline of the proofs of the main theorems is given in Section 6. Supplementary Material section gives a description for a supplemental document by Chen, Gao and Li [5], from which the detailed proofs of the main theorems, along with some technical lemmas, are available.

2. Estimation method and assumptions

2.1. Markov theory

Let $\{V_t, t \geq 0\}$ be a Markov chain with transition probability P and state space $(\mathbf{E}, \mathcal{E})$, and ϕ be a measure on $(\mathbf{E}, \mathcal{E})$. Throughout the paper, $\{V_t\}$ is assumed to be ϕ -irreducible Harris recurrent, which makes asymptotics for semi-parametric estimation possible. The class of stochastic processes we are dealing with in this paper is not the general class of null recurrent Markov chains. Instead, we need to impose some restrictions on the tail behavior of the distribution of the recurrence time S_α of the chain. This is what we are interested in: a class of β null recurrent Markov chains.

Definition. A Markov chain $\{V_t\}$ is β null recurrent if there exist a small non-negative function $f(\cdot)$ (the definition of a small function can be found in the supplemental document), an initial measure λ , a constant $\beta \in (0, 1)$ and a slowly varying function $L_f(\cdot)$ such that

$$\mathbf{E}_\lambda \left[\sum_{t=0}^n f(V_t) \right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n) \quad \text{as } n \rightarrow \infty, \quad (2.1)$$

where \mathbf{E}_λ stands for the expectation with initial distribution λ and $\Gamma(\cdot)$ is the usual gamma function.

It is shown in Karlsen and Tjøstheim [20] that when there exist some small measure ν and small function s with $\nu(\mathbf{E}) = 1$ and $0 \leq s(v) \leq 1$, $v \in \mathbf{E}$, such that

$$P \geq s \otimes \nu, \quad (2.2)$$

then $\{V_t\}$ is β null recurrent if and only if

$$\mathbf{P}_\alpha(S_\alpha > n) = \frac{1}{\Gamma(1 - \beta)n^\beta L_s(n)}(1 + o(1)), \quad (2.3)$$

where $L_s = \frac{L_f}{\pi_s f}$ and π_s is the invariant measure as defined in Karlsen and Tjøstheim [20]. Furthermore, if (2.3) holds, by Lemma 3.4 in Karlsen and Tjøstheim [20], $\hat{\beta} := \frac{\ln(N_{\mathbf{C}}(n))}{\ln n}$ is a strongly consistent estimator of β , where $N_{\mathbf{C}}(n) = \sum_{t=1}^n I_{\mathbf{C}}(V_t)$, in which $I_A(\cdot)$ is the conventional indicator function and \mathbf{C} is a small set as defined in Karlsen and Tjøstheim [20].

We then introduce a useful decomposition that is critical in the proofs of asymptotics for nonparametric estimation in null recurrent time series. Let f be a real function defined in \mathbf{R} . We now decompose the partial sum $S_n(f) = \sum_{t=0}^n f(V_t)$ into a sum of independent and identically distributed (i.i.d.) random variables with one main part and two asymptotically negligible minor parts. Define

$$Z_k = \begin{cases} \sum_{t=0}^{\tau_0} f(V_t), & k = 0, \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} f(V_t), & 1 \leq k \leq N(n), \\ \sum_{t=\tau_{N(n)}+1}^n f(V_t), & k = (n), \end{cases}$$

where the definitions of τ_k and $N(n)$ will be given in the supplemental document. Then

$$S_n(f) = Z_0 + \sum_{k=1}^{N(n)} Z_k + Z_{(n)}. \quad (2.4)$$

From Nummelin's [24] result, we know that $\{Z_k, k \geq 1\}$ is a sequence of i.i.d. random variables. In the decomposition (2.4) of $S_n(f)$, $N(n)$ plays the role of the number of observations. It follows from Lemma 3.2 in Karlsen and Tjøstheim [20] that Z_0 and $Z_{(n)}$ converge to zero almost surely when they are divided by $N(n)$. Furthermore, Karlsen and Tjøstheim [20] show that if (2.2) holds and $\int |f(v)|\pi_s(dv) < \infty$, then for an arbitrary initial distribution λ we have

$$\frac{1}{N(n)} S_n(f) \longrightarrow \pi_s(f) \quad \text{almost surely (a.s.)}, \quad (2.5)$$

where $\pi_s(f) = \int f(v)\pi_s(dv)$.

Some useful results for Markov theory are available from Appendix A of the supplemental document.

2.2. Estimation method

As assumed in assumption A1 below, there exist a function $H(\cdot)$ and a stationary process $\{U_t\}$ such that $X_t = H(V_t) + U_t$. Since $E[\epsilon_t|V_t = v] = E[\epsilon_t] = 0$ is assumed in A2(ii) and A3(ii), we have

$$E[Y_t|V_t = v] = E[(X_t^\tau \theta_0 + g(V_t) + \epsilon_t)|V_t = v] = H(v)^\tau \theta_0 + g(v). \quad (2.6)$$

This implies that $\Psi(v) \equiv E[Y_t|V_t = v]$ is a function of v independent of t for each fixed v and given θ_0 . Thus, the form of $g(v)$ can be represented by

$$g(v) = \Psi(v) - H(v)^\tau \theta_0. \quad (2.7)$$

In view of (2.7), we can rewrite model (1.3) as

$$Y_t - \Psi(V_t) = (X_t - H(V_t))^\tau \theta_0 + \epsilon_t. \quad (2.8)$$

Letting $W_t = Y_t - \Psi(V_t)$ and $U_t = X_t - H(V_t)$, model (2.8) implies

$$W_t = Y_t - \Psi(V_t) = (X_t - H(V_t))^\tau \theta_0 + \epsilon_t = U_t^\tau \theta_0 + \epsilon_t. \quad (2.9)$$

Note that $E[W_t] = E[U_t^\tau \theta_0] + E[\epsilon_t] = 0$. In the case where $\{(X_t, V_t, \epsilon_t): t \geq 1\}$ is a sequence of stationary random variables, various estimation methods for θ_0 and $g(\cdot)$ in model (1.3) have been studied by many authors (see, e.g., Robinson [27], Härdle *et al.* [15] and Gao [10]).

We now propose an SLS estimation method based on the kernel smoothing. For every given θ , we define a kernel estimator of $g(v)$ by

$$g_n(v; \theta) = \sum_{t=1}^n w_{nt}(v) (Y_t - X_t^\tau \theta), \quad (2.10)$$

where $\{w_{nt}(v)\}$ is a sequence of weight functions given by

$$w_{nt}(v) = \frac{K_{v,h}(V_t)}{\sum_{k=1}^n K_{v,h}(V_k)} \quad \text{with } K_{v,h}(V_t) = \frac{1}{h} K\left(\frac{V_t - v}{h}\right),$$

in which $K(\cdot)$ is a probability kernel function and $h = h_n$ is a bandwidth parameter.

Replacing $g(V_t)$ by $g_n(V_t; \theta)$ in model (1.3) and applying the SLS estimation method, we obtain the SLS estimator, $\bar{\theta}_n$, of θ_0 by minimizing

$$\frac{1}{n} \sum_{t=1}^n (Y_t - X_t^\tau \theta - g_n(V_t, \theta))^2$$

over θ . This implies

$$\bar{\theta}_n = (\bar{X}^\tau \bar{X})^{-1} \bar{X}^\tau \bar{Y}, \quad (2.11)$$

where $\overline{X}^\tau = (\tilde{X}_1, \dots, \tilde{X}_n)$, $\tilde{X}_t = X_t - \sum_{k=1}^n w_{nk}(V_t)X_k$, $\overline{Y}^\tau = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ and $\tilde{Y}_t = Y_t - \sum_{k=1}^n w_{nk}(V_t)Y_k$. And $g(\cdot)$ is then estimated by

$$\overline{g}_n(\cdot) = g_n(\cdot; \overline{\theta}_n). \quad (2.12)$$

This kind of estimation method has been studied in the literature (see, e.g., Härdle *et al.* [15]). When $\{V_t\}$ is a sequence of either fixed designs or stationary regressors with a compact support, the conventional weighted least-squares estimators (2.11) and (2.12) work well in both the large and small sample cases. Since the invariant distribution of β null recurrent Markov chain $\{V_t\}$ might not have any compact support, it is difficult to establish asymptotic results for the estimators (2.11) and (2.12) owing to the random denominator problem involved in $w_{nt}(\cdot)$. Hence, to establish our asymptotic theory, we apply the following weighted least-squares estimation method (see, e.g., Robinson [27]). Define

$$F_t := F_{nt} = I(|p_n(V_t)| > b_n), \quad (2.13)$$

where

$$p_n(v) = \frac{1}{N(n)} \sum_{k=1}^n K_{v,h}(V_k)$$

and $\{b_n\}$ is a sequence of positive numbers satisfying some conditions. Furthermore, let

$$\tilde{X}^\tau = (\tilde{X}_1 F_1, \dots, \tilde{X}_n F_n) \quad \text{and} \quad \tilde{Y}^\tau = (\tilde{Y}_1 F_1, \dots, \tilde{Y}_n F_n).$$

Throughout this paper, we propose to estimate θ_0 by

$$\hat{\theta}_n = (\tilde{X}^\tau \tilde{X})^{-1} \tilde{X}^\tau \tilde{Y} \quad (2.14)$$

and $g(\cdot)$ by

$$\hat{g}_n(\cdot) = g_n(\cdot; \hat{\theta}_n). \quad (2.15)$$

2.3. Assumptions

As may be seen from equation (2.9), further discussion on the semi-parametric estimation method depends heavily on the structure of $\{X_t\}$ and $\{V_t\}$. This paper is concerned with the following two cases: (i) where $\{X_t\}$ is a sequence of strictly stationary regressors and independent of $\{V_t\}$; and (ii) where $\{X_t\}$ is a sequence of non-stationary regressors with the non-stationarity being generated by $\{V_t\}$.

Before stating the main assumptions, we introduce the definition of α mixing dependence. The stationary sequence $\{Z_t, t = 0, \pm 1, \dots\}$ is said to be α mixing if $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$, where

$$\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |P(AB) - P(A)P(B)|,$$

in which $\{\mathcal{F}_k^j\}$ denotes a sequence of σ fields generated by $\{Z_t, k \leq t \leq j\}$. Since its introduction by Rosenblatt [29], α mixing dependence is a property shared by many time series models (see, e.g., Withers [34] and Gao [10]). For more details about limit theorems for α mixing processes, we refer to Lin and Lu [22] and the references therein.

The following assumptions are necessary to derive the asymptotic properties of the semi-parametric estimators.

A1. There exist an unknown function $H(v)$ and a stationary process $\{U_t\}$ such that $X_t = H(V_t) + U_t$.

A2. (i) Suppose that $\{U_t\}$ is a stationary ergodic Markov process with $E[U_1] = 0$ and $E[\|U_1\|^{4+\gamma_1}] < \infty$ for some $\gamma_1 > 0$, where $\|\cdot\|$ stands for the Euclidean norm. Furthermore, we suppose that $\Sigma := E[U_1 U_1^T]$ is positive definite and $\{U_t\}$ is α mixing with

$$\sum_{t=1}^{\infty} \alpha_U^{\gamma_1/(4+\gamma_1)}(t) < \infty, \quad (2.16)$$

where $\alpha_U(t)$ is the α mixing coefficient of $\{U_t\}$.

(ii) Let $\{\epsilon_t\}$ be a stationary ergodic Markov process with $E[\epsilon_1] = 0$, $\sigma^2 := E[\epsilon_1^2] > 0$ and $E[|\epsilon_1|^{2+\gamma_2}] < \infty$ for some $\gamma_2 > 0$. Furthermore, the process $\{\epsilon_t\}$ is α mixing with

$$\sum_{t=1}^{\infty} \alpha_{\epsilon}^{\gamma_2/(2+\gamma_2)}(t) < \infty, \quad (2.17)$$

where $\alpha_{\epsilon}(t)$ is the α mixing coefficient of $\{\epsilon_t\}$.

A3. (i) The invariant measure π_s of the β null recurrent Markov chain $\{V_t\}$ has a uniformly continuous density function $p_s(\cdot)$.

(ii) Let $\{U_t\}$, $\{V_t\}$ and $\{\epsilon_t\}$ be mutually independent.

A4. Let $f_{i,k}(\cdot)$ be the density function of

$$V_{i,k} = \varphi_{i-k}(V_i - V_k) \quad \text{for } i > k \text{ with } \varphi_m = m^{\beta-1} L_s(m) \text{ for } m \geq 1.$$

Let

$$\inf_{\delta > 0} \limsup_{m \rightarrow \infty} \sup_{i \geq 1} \sup_{|v| \leq \delta} f_{i+m,i}(v) < \infty. \quad (2.18)$$

Furthermore, there exists a sequence of σ fields $\{\mathcal{F}_t, t \geq 0\}$ such that $\{V_t\}$ is adapted to \mathcal{F}_t . With probability 1,

$$\inf_{\delta > 0} \limsup_{m \rightarrow \infty} \sup_{i \geq 1} \sup_{|v| \leq \delta} f_{i+m,i}(v | \mathcal{F}_i) < \infty, \quad (2.19)$$

where $f_{i,k}(v | \mathcal{F}_k)$ is the conditional density function of $V_{i,k}$ given \mathcal{F}_k .

A5. (i) The function $g(v)$ is differentiable and the derivative is continuous in $v \in \mathbf{R}$. In addition, for n large enough

$$\sum_{t=1}^n \int (g'(\varphi_t^{-1}v))^2 f_{t,0}(v) dv = O(nh^{-1}), \quad (2.20)$$

where $g'(\cdot)$ is the derivative of $g(\cdot)$, the definitions of φ_t and $f_{t,0}(v)$ are given in A4 above.

(ii) The function $H(v)$ is differentiable and the derivative is also continuous in $v \in \mathbf{R}$. In addition, for n large enough

$$\sum_{t=1}^n \int \|H'(\varphi_t^{-1}v)\|^2 f_{t,0}(v) dv = O(nh^{-1}) \quad (2.21)$$

and

$$\sum_{t=1}^n \int \|g'(\varphi_t^{-1}v)H'(\varphi_t^{-1}v)\| f_{t,0}(v) dv = O(n^{1/2-\varepsilon_1}b_n^2h^{-2}), \quad (2.22)$$

where $\varepsilon_1 > 0$ is small enough.

A6. (i) The probability kernel function $K(\cdot)$ is a continuous and symmetric function having some compact support.

(ii) The sequences $\{h_n\}$ and $\{b_n\}$ both satisfy as $n \rightarrow \infty$

$$h_n \rightarrow 0, \quad b_n \rightarrow 0, \quad n^{\varepsilon_0} h_n b_n^{-4} \rightarrow 0 \quad \text{and} \quad n^{\beta-\varepsilon_0} h_n b_n^4 \rightarrow \infty \quad (2.23)$$

for some $0 < \varepsilon_0 < \frac{\beta}{2}$. Moreover,

$$\sum_{t=1}^n P(p_n(V_t) \leq b_n) = o(n). \quad (2.24)$$

Remark 2.1. (i) While some parts of assumptions A1–A3 may be non-standard, they are justifiable in many situations. Condition A1 assumes that $\{X_t\}$ is generated by $X_t = H(V_t) + U_t$. This is satisfied when the conditional mean function $H(v) = E[X_t|V_t = v]$ exists. In this case, A1 holds automatically with $U_t = X_t - E[X_t|V_t]$. Condition A1 is also commonly used in the stationary case (see, e.g., Linton [23]). There are various examples in this kind of situation (see, e.g., in the univariate case where $X_t = V_t + \varepsilon_t$, in which $\{\varepsilon_t\}$ is a sequence of i.i.d. errors with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] < \infty$, and independent of $\{V_t\}$. In this case, $H(v) = E[X_t|V_t = v] = v$ and $U_t = \varepsilon_t$). As a consequence, condition A1 does not include the case where $\{X_t\}$ is a random walk sequence of the form $X_t = X_{t-1} + \zeta_t$. Note that the case where the non-stationarity in both $\{X_t\}$ and $\{V_t\}$ is generated by a common random walk structure will need to be discussed separately, since the methodology involved is likely to be quite different. In Section 3.2 below, we will give some discussion about the case where $\{H(V_t)\}$ is replaced by a bivariate function of the form $\{H(V_t, t)\}$ to take into account the inhomogeneous case.

(ii) The stationarity assumption on $\{U_t\}$ is to ensure that the conventional \sqrt{n} -rate of convergence is achievable and thus it is possible to construct an asymptotically efficient estimator for θ_0 . The stationarity condition on $\{U_t\}$ also requires that X_t can be decomposed into a non-stationary component represented by $H(V_t)$ and a stationary component $\{U_t\}$. The α mixing dependence in A2 is a mild condition on $\{U_t\}$ and the

errors process $\{e_t\}$. Karlsen *et al.* [19] have made similar assumptions. As discussed in Section 3.2 below, A2(i) can be relaxed to allow for the inclusion of both endogeneity and heteroscedasticity. Note that A2(ii) can also be relaxed to allow for the inclusion of a deterministic function in model (1.3). In such cases, model (1.3) can be naturally extended to a semi-parametric additive model of the form $Y_t = X_t^\tau \theta_0 + g(V_t) + \lambda(U_t, t) + e_t$ as discussed in Section 3.2 below.

(iii) As we can see from the asymptotic theory below, the condition on the existence of the inverse matrix Σ^{-1} is required in Theorem 3.1. In the case where $\{(X_t, V_t)\}$ is a vector of either independent regressors or stationary time series regressors, Härdle *et al.* [15] also assume similar conditions (see Section 1.3 in their book) for establishing the asymptotic results for the conventional least-squares estimators of θ_0 in (2.11) and of $g(\cdot)$ in (2.12). Condition A3(i) corresponds to analogous conditions on the density function in the stationary case. A3(ii) imposes the mutual independence to avoid involving some extremely technical conditions.

Remark 2.2. A4 is similar to but weaker than Assumption 2.3(ii) in Wang and Phillips [32]. It is easy to check that (2.18) and (2.19) are satisfied with $\beta = 1$ and $L_s(\cdot) \equiv 1$ when $\{V_t\}$ is a sequence of either i.i.d. or stationary dependent variables. Consider the random walk case defined by

$$V_t = V_{t-1} + v_t, \quad t = 1, 2, \dots, V_0 = 0, \quad (2.25)$$

where $\{v_t\}$ is a sequence of i.i.d. random variables. The random walk model (2.25) is very important in economics and finance and has been studied by many authors. It corresponds to a $1/2$ null recurrent process and it is easy to check that (2.18) and (2.19) are satisfied with $\beta = 1/2$, $L_s(n) \equiv 1$ and $\mathcal{F}_k = \sigma(v_i, i \leq k)$. On the other hand, (2.18) and (2.19) can be formulated in terms of the transition probability. For example, assume that the transition probability of the Markov process $\{V_t\}$ is defined by

$$P(x, dy) = f(x|y) dy.$$

Let $f^k(\cdot)$ be the marginal density of $\{V_k\}$ and $f^m(x|y)$ be the m step transition density. Then

$$f_{i+m,i}(v) = \varphi_m^{-1} \int f^m(\varphi_m^{-1}v + y|y) f^i(y) dy,$$

where φ_m is defined in A4.

Remark 2.3. (i) A5(i) is assumed to make sure that the bias term of the nonparametric estimator is negligible when establishing the asymptotic distribution of the semi-parametric estimator θ_n . When $\{V_t\}$ is the random walk process defined by (2.25), condition A5(i) can be verified. If

$$g(v) = \varrho_0 + \varrho_1 v + \varrho_2 |v|^{1+\delta_0}, \quad 0 < \delta_0 < 1/2, \quad (2.26)$$

$n^{\delta_0}h = O(1)$ and $f_{t,0}(v) = O(v^{-(1+2\delta_0+\varsigma)})$ for some $\varsigma > 0$ as $t \rightarrow \infty$ and $v \rightarrow \infty$, we can show that by A4,

$$\sum_{t=1}^n \int (g'(\varphi_t^{-1}v))^2 f_{t,0}(v) dv = O\left(\sum_{t=1}^n \varphi_t^{-2\delta_0}\right) = O(n^{1+\delta_0}),$$

which implies (2.20).

(ii) Similarly, condition A5(ii) is also verifiable. Consider the case where

$$g(v) = \varrho_0 + \varrho_1 v \quad \text{and} \quad H(v) = \mathbf{a}_0 + \mathbf{a}_1 v + \mathbf{a}_2 |v|^{1+\delta_1}, \quad 1 < \delta_1 < 1/2,$$

in which \mathbf{a}_k , $k = 0, 1, 2$, are d -dimensional vectors, $n^{1/2+\delta_1-\varepsilon_1}h^2 = O(1)$ ($\varepsilon_1 < \frac{1}{2} - \delta_1$) and $f_{t,0}(v) = O(v^{-(1+2\delta_1+\varsigma)})$ for some $\varsigma > 0$ as $t \rightarrow \infty$ and $v \rightarrow \infty$. We can also show that (2.21) and (2.22) hold for the random walk case. The detailed calculation is similar to that in Remark 2.3(i) above.

Remark 2.4. (i) Condition A6(i) is a quite natural condition on the kernel function and has been used by many authors for the stationary time series case. The first part of A6(i) requires that the rate of $b_n^{-4} \rightarrow \infty$ is slower than that of $n^{\varepsilon_0}h \rightarrow 0$ and the rate of $b_n^4 \rightarrow 0$ is slower than that of $n^{\beta-\varepsilon_0}h \rightarrow \infty$. Such conditions are satisfied in various cases. Letting $b_n = c_b \log^{-1}(n)$ and $h_n = c_h n^{-\zeta_0}$ for some $c_b > 0$, $c_h > 0$ and $\varepsilon_0 < \zeta_0 < \beta - \varepsilon_0$, then the first part of A6(ii) holds automatically.

(ii) The second part of A6(ii) is imposed to ensure that the truncated procedure works in this kind of problem. When $\{V_t\}$ is a sequence of i.i.d. random variables having some compact support S , it is easy to show that (2.24) holds if $\inf_{x \in S} p(x) > 0$, where $p(\cdot)$ is the density function of $\{V_t\}$. In the case where $\{V_t\}$ is an i.i.d. sequence without any compact support, Robinson [27] gives different conditions such that (2.24) holds. We can show that condition A6(ii) is verifiable when $\{V_t\}$ is a random walk model of the form (2.25). Since the verification is quite technical, the details are given in the last part of Appendix C in the supplemental document.

3. The main results and their extensions

3.1. Asymptotic theory

We now establish an asymptotic distribution of the estimate $\hat{\theta}_n$ in the following theorem. The following theorem includes two cases: (a) $\{V_t\}$ is a sequence of non-stationary regressors and $\{X_t\}$ is a sequence of strictly stationary regressors and is independent of $\{V_t\}$; and (b) both $\{X_t\}$ and $\{V_t\}$ are non-stationary.

Theorem 3.1. *Let A1–A5(i) and A6 hold. In addition, suppose that $\Sigma_{\epsilon,U} := \sigma^2 \Sigma + 2 \sum_{t=2}^{\infty} E[\epsilon_1 \epsilon_t] E[U_1 U_t^T]$ is positive definite.*

(i) If $\{X_t\}$ is strictly stationary and independent of $\{V_t\}$, then as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1} \Sigma_{\epsilon, U} \Sigma^{-1}). \quad (3.1)$$

(ii) Suppose that both $\{X_t\}$ and $\{V_t\}$ are non-stationary. If, in addition, A5(ii) is satisfied, then (3.1) still holds.

Remark 3.1. (i) Theorem 3.1 shows that the standard normality can still be an asymptotic distribution of the SLS estimate even when non-stationarity is involved. Theorem 3.1(ii) further shows that the conventional rate of \sqrt{n} is still achievable when the non-stationarity in $\{X_t\}$ is purely generated by $\{V_t\}$ and certain conditions are imposed on the functional forms of $H(\cdot)$ and $g(\cdot)$.

(ii) Since the asymptotic distribution and asymptotic variance in (3.1) are mainly determined by the stationary sequences $\{\epsilon_t\}$ and $\{U_t\}$, the above conclusion extends Theorem 2.1.1 of Härdle *et al.* [15] for the case when $\{X_t\}$, $\{V_t\}$ and $\{\epsilon_t\}$ are all strictly stationary. In addition, when $\{X_t\}$ is assumed to be strictly stationary and independent of $\{V_t\}$ in Theorem 3.1(i), the covariance matrix reduces to the covariance matrix of $\{X_t\}$ of the form $\Sigma = E[(X_1 - E[X_1])(X_1 - E[X_1])^T]$.

Remark 3.2. (i) Theorem 3.1 establishes an asymptotically normal estimator for θ_0 . As in the independent and stationary sample case, an interesting issue is how to construct an asymptotically efficient estimator for θ_0 . As discussed in Chen [4] and Härdle *et al.* [15], it can be shown that $\hat{\theta}$ achieves the smallest possible variance of $\sigma^2 \Sigma^{-1}$ when both $\{U_t\}$ and $\{\epsilon_t\}$ are independent and $\epsilon_t \sim N(0, \sigma^2)$.

(ii) Since the publication of the book by Bickel *et al.* [3], there has been an increasing interest in the field of asymptotic efficiency in semi-parametric models. There are certain types of asymptotic efficiency in this kind of semi-parametric setting. Härdle *et al.* [15] consider several types of asymptotically efficient estimators in Chapters 2 and 5 of the book. Linton [23] considers second-order efficiency. Bhattacharya and Zhao [2] establish an asymptotically efficient estimator without requiring finite variance. Chen [6] discusses asymptotic efficiency in nonparametric and semi-parametric models using sieve estimation.

(iii) As shown in the literature, the establishment of an asymptotically efficient estimator in this kind of semi-parametric setting requires the availability of uniform convergence of nonparametric estimation. Since such uniform convergence results are not readily available and applicable in this kind of non-stationary situation, we wish to establish some necessary uniform convergence results first before we may be able to address the issue of asymptotic efficiency in future research.

An asymptotic distribution of $\hat{g}_n(x)$ is given in Theorem 3.2 below.

Theorem 3.2. (i) Let the conditions of Theorem 3.1(i) hold. If, in addition, $g(\cdot)$ is twice differentiable and the second derivative, $g''(v)$, is continuous in v and $n^{\beta/5+\varepsilon}h = o(1)$ for

some $\varepsilon > 0$, then as $n \rightarrow \infty$,

$$\sqrt{\sum_{t=1}^n K\left(\frac{V_t - v}{h}\right)} (\hat{g}_n(v) - g(v)) \xrightarrow{d} N\left(0, \sigma^2 \int K^2(u) du\right). \quad (3.2)$$

(ii) Let the conditions of Theorem 3.1(ii) hold. If, in addition, $g(\cdot)$ is twice differentiable and the second derivative, $g''(v)$, is continuous in v and $n^{\beta/5+\varepsilon}h = o(1)$ for some $\varepsilon > 0$, then equation (3.2) remains true.

Remark 3.3. The asymptotic distribution in (3.2) is similar to the corresponding results obtained by Karlsen *et al.* [19] and Wang and Phillips [32]. The rate of convergence is slower than that for the stationary time series case as $\sum_{t=1}^n K\left(\frac{V_t - v}{h}\right) = O_P(N(n)h)$ and $N(n)$ is usually smaller than n almost surely. The condition $n^{\beta/5+\varepsilon}h = o(1)$ makes sure that the bias term of the nonparametric estimator $\hat{g}_n(v)$ is negligible.

3.2. Some extensions

In this section, we give some detailed discussion of the possible extensions raised in Remark 2.1(ii) and (iii). In addition, we also suggest some other extensions.

Instead of considering a variety of extensions of model (1.3) and Theorems 3.1 and 3.2, this section considers several extensions that are naturally based on the relaxation of A1–A3 to Assumptions 3.1–3.3 below, respectively. As a consequence, the extended models proposed below allow for the inclusion of endogeneity, heteroscedasticity and deterministic trending.

Assumption 3.1. There are a bivariate function $H(\cdot, \cdot)$ and a stationary process $\{U_t\}$ such that $X_t = H(V_t, \frac{t}{n}) + U_t$ for $1 \leq t \leq n$.

Assumption 3.2. (i) Let A2(i) hold.

(ii) Let $\{\epsilon_t\}$ be of the form of either $\epsilon_t = \sigma(\zeta_t)e_t$ or $\epsilon_t = \lambda(\xi_t) + e_t$ with $\zeta_t = U_t$ or V_t and $\xi_t = U_t$ or $\xi_t = \frac{t}{n}$, in which $\{e_t\}$ is a stationary ergodic Markov process satisfying A2(ii) and both $\sigma(\cdot)$ and $\lambda(\cdot)$ are smooth functions.

Assumption 3.3. (i) Let A3(i) hold.

(ii) Let $\{V_t\}$ be independent of both $\{U_t\}$ and $\{e_t\}$. In addition, $E[e_t|U_t] = 0$.

While it is difficult to consider some general non-stationarity for $\{X_t\}$, it is possible to consider a general inhomogeneous case in Assumption 3.1 to allow for a bivariate functional form of $H(\cdot, \cdot)$ such that the non-stationarity of $\{X_t\}$ is caused by both the involvement of $\{V_t\}$ and the dependence on t . In this case, $H(\cdot, \cdot)$ may be estimated nonparametrically by

$$\hat{H}(v, \tau) = \sum_{t=1}^n W_{nt}(v, \tau) X_t \quad \text{with } W_{nt}(v, \tau) = \frac{K_{v,\tau}(V_t, t)}{\sum_{k=1}^n K_{v,\tau}(V_k, k)}, \quad (3.3)$$

where $K_{v,\tau}(V_t, t) = \frac{1}{h_1} \frac{1}{h_2} K_1(\frac{V_t - v}{h_1}) K_2(\frac{t/n - \tau}{h_2})$, in which both $K_i(\cdot)$ are probability kernel functions and h_i are bandwidth parameters for $i = 1, 2$.

Assumption 3.2(ii) allows for inclusion of endogeneity, heteroscedasticity and deterministic trending. In the case where we have either $\epsilon_t = \sigma(U_t)e_t$ or $\epsilon_t = \sigma(V_t)e_t$ with $E[e_t|U_t] = E[e_t|V_t] = 0$, it follows that either $E[\epsilon_t|V_t] = E[\sigma(V_t)e_t|V_t] = \sigma(V_t)E[e_t|V_t] = 0 = E[\epsilon_t]$ or $E[\epsilon_t|V_t] = E[\sigma(U_t)e_t] = E[\epsilon_t]$. This implies Assumption 3.2(ii) holds in both cases. In addition, Assumption 3.2(ii) also includes the case where $\epsilon_t = \lambda(\frac{t}{n}) + e_t$ or $\epsilon_t = \lambda(U_t) + e_t$. In such cases, obviously we have $E[\epsilon_t|V_t] = E[\epsilon_t]$.

Under Assumptions 3.1–3.3, model (1.3) can be written as either

$$\begin{aligned} Y_t &= X_t^\tau \theta_0 + g(V_t) + \sigma(\zeta_t)e_t, \\ X_t &= H\left(V_t, \frac{t}{n}\right) + U_t, \end{aligned} \tag{3.4}$$

where $\zeta_t = U_t$ or V_t , or

$$\begin{aligned} Y_t &= X_t^\tau \theta_0 + g(V_t) + \lambda(\xi_t) + e_t, \\ X_t &= H\left(V_t, \frac{t}{n}\right) + U_t, \end{aligned} \tag{3.5}$$

where $\xi_t = U_t$ or $\xi_t = \frac{t}{n}$.

Estimation of θ_0 and $g(\cdot)$ in (3.4) is similar to what has been proposed in Section 2. Since model (3.5) is a semi-parametric additive model, one will need to estimate θ_0 based on the form $Y_t = X_t^\tau \theta_0 + G(V_t, \xi_t) + e_t$ with $G(v, \tau) = g(v) + \lambda(\tau)$ before both $g(\cdot)$ and $\lambda(\cdot)$ can be individually estimated using the marginal integration method as developed in Section 2.3 of Gao [10].

In both cases, one will need to replace $\{w_{nt}(v)\}$ in (2.10) and $p_n(v)$ in (2.13) by $\{W_{nt}(v, \tau)\}$ of (3.3) and $p_n(v, \tau) = \frac{1}{N(n)} \sum_{k=1}^n K_{v,\tau}(V_k, k)$, respectively.

Since the establishment and the proofs of the corresponding results of Theorems 3.1 and 3.2 for models (3.4) and (3.5) involve more technicalities than those given in Appendices B and C of the supplemental document, we wish to leave the discussion of models (3.4) and (3.5) to a future paper.

4. Simulation study

To illustrate our estimation procedure, we consider a simulated example and a real data example in this section. Throughout the section, the uniform kernel $K(v) = \frac{1}{2}I_{[-1,1]}(v)$ is used. A difficult problem in simulation is the choice of a proper bandwidth. From the asymptotic results in Section 3, we can find that the rates of convergence are different from those in the stationary case with n being replaced by $N(n)$. In practice, we have found it useful to use a semi-parametric cross-validation method (see, e.g., Section 2.1.3 of Härdle *et al.* [15]).

Table 1. Simulation results for the estimator of θ_0

n	$H(\cdot)$	AE	SE
200	$H(v) \equiv 0$	0.0137	0.0144
700	$H(v) \equiv 0$	0.0117	0.0086
1200	$H(v) \equiv 0$	0.0064	0.0062
200	$H(v) = v$	0.0172	0.0215
700	$H(v) = v$	0.0149	0.0126
1200	$H(v) = v$	0.0079	0.0108

Example 4.1. Consider a partially linear time series model of the form

$$Y_t = X_t\theta + g(V_t) + \epsilon_t, \quad t = 1, 2, \dots, n, \quad (4.1)$$

where $V_t = V_{t-1} + v_t$ with $V_0 = 0$ and $\{v_t\}$ is a sequence of i.i.d. random variables generated from $N(0, 0.1^2)$, $\{\epsilon_t\}$ is generated by an AR(1) model of the form

$$\epsilon_t = 0.5\epsilon_{t-1} + \eta_t,$$

in which $\{\eta_t\}$ is a sequence of i.i.d. random variables generated from $N(0, 1)$, $\{v_t\}$ and $\{\eta_t\}$ are mutually independent. We then choose the true value of θ as $\theta_0 = 1$, the true form of $g(\cdot)$ as $g_0(v) = v$ and consider the following cases for $\{X_t\}$.

- (i) $X_t = U_t$, where $\{U_t\}$ is a sequence of i.i.d. $N(0, 1)$ random variables,
- (ii) $X_t = V_t + U_t$, where $\{U_t\}$ is defined as in case (i).

It is easy to check that the random walk $\{V_t\}$ defined in this example corresponds to a 1/2 null recurrent process and the assumptions in Section 2 are satisfied here. We choose sample sizes $n = 200, 700, 1200$ and $N = 1000$ as the number of replications in the simulation. The simulation results are listed in Tables 1 and 2 and the plots are given in Figures 1–6.

The performance of $\hat{\theta}_n$ is given in Table 1. The “AE” in Table 1 is defined by $\frac{1}{1000} \sum_{j=1}^{1000} |\hat{\theta}(j) - \theta_0|$, where $\hat{\theta}(j)$ is the value of $\hat{\theta}_n$ in the j th replication. “SE” is

Table 2. Simulation results for the estimator of $g_0(v) = v$

n	$H(\cdot)$	AE	SE
200	$H(v) \equiv 0$	0.1158	0.0575
700	$H(v) \equiv 0$	0.0894	0.0341
1200	$H(v) \equiv 0$	0.0628	0.0210
200	$H(v) = v$	0.1391	0.0582
700	$H(v) = v$	0.1299	0.0437
1200	$H(v) = v$	0.1075	0.0367

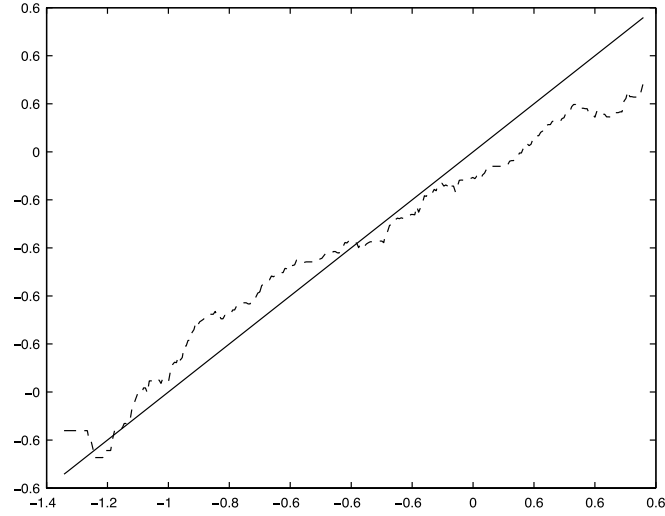


Figure 1. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) \equiv 0$ with sample size $n = 200$; the solid line is the true line, and the dashed curve is the estimated curve.

the standard error of $\{\hat{\theta}(j)\}$. From Table 1, we find that the estimator of θ_0 performs well in the small and medium sample cases and it improves when the sample size increases.

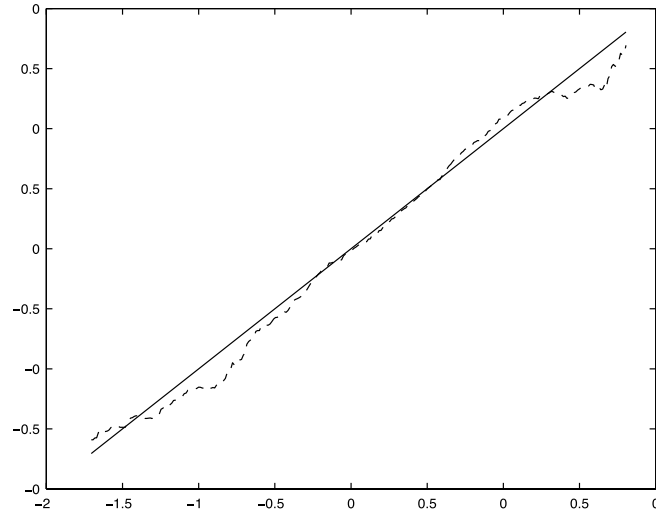


Figure 2. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) \equiv 0$ with sample size $n = 700$; the solid line is the true line, and the dashed curve is the estimated curve.

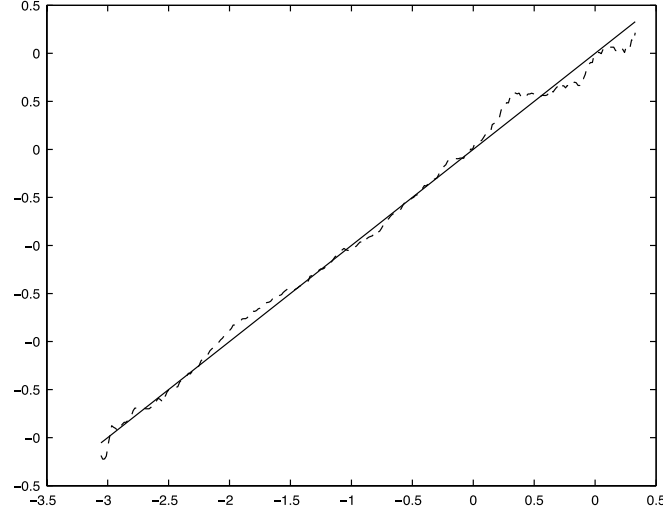


Figure 3. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) \equiv 0$ with sample size $n = 1200$; the solid line is the true line, and the dashed curve is the estimated curve.

The performance of the nonparametric estimator is given in Table 2. The “AE” in Table 2 is the mean of the absolute errors in 1000 replications. The absolute error is defined by $\frac{1}{300} \sum_{j=1}^{300} |\hat{g}_n(v_j) - v_j|$, where $v_j = v_{\min} + \frac{j-1}{300}(v_{\max} - v_{\min})$ for $j = 1, 2, \dots, 300$, v_{\max}

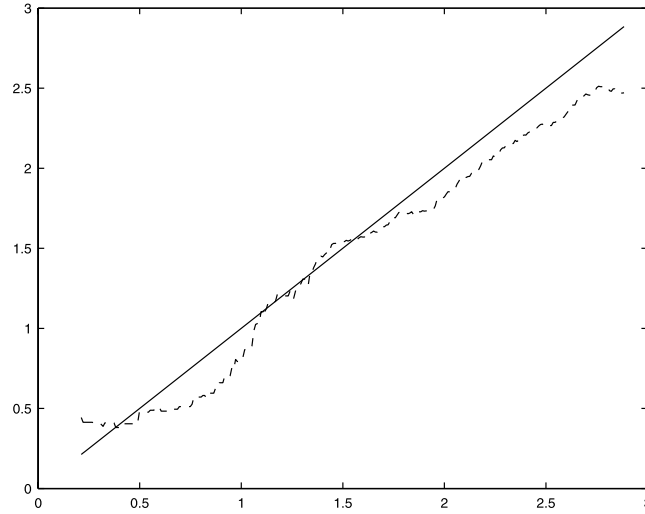


Figure 4. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) = v$ with sample size $n = 200$; the solid line is the true line, and the dashed curve is the estimated curve.

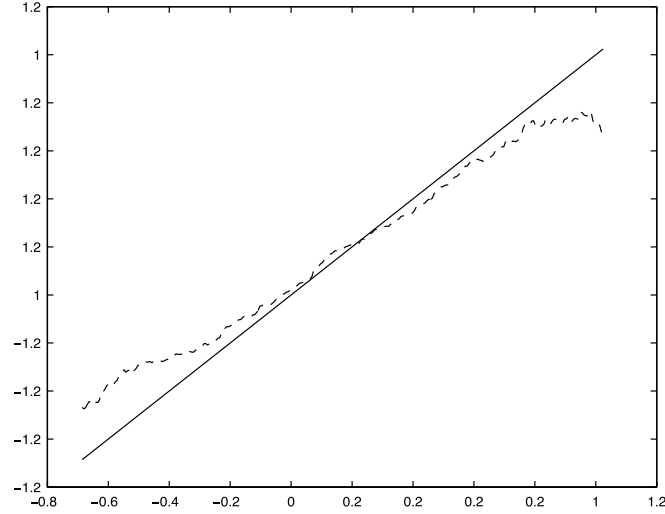


Figure 5. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) = v$ with sample size $n = 700$; the solid line is the true line, and the dashed curve is the estimated curve.

and v_{\min} are the maximum and minimum of the random walk $\{V_t, 1 \leq t \leq n\}$, respectively. “SE” in Table 2 is the standard error. From Table 2, we find that the nonparametric estimate of $g_0(v) = v$ performs well in our example and it improves when the sample size increases.

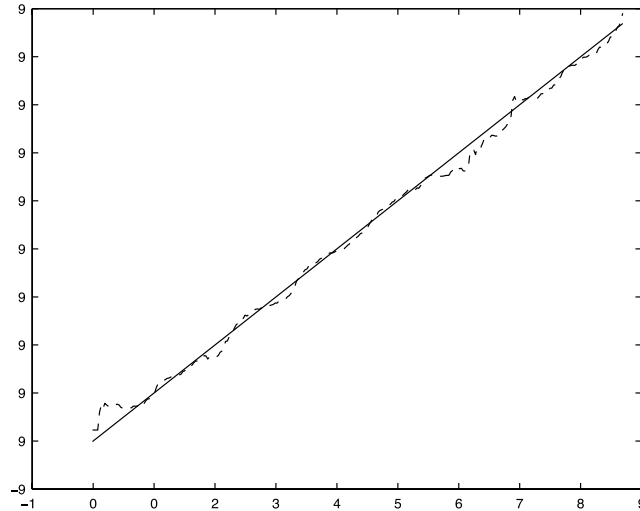


Figure 6. Nonparametric estimate of the regression function $g_0(v)$ for the case of $H(v) = v$ with sample size $n = 1200$; the solid line is the true line, and the dashed curve is the estimated curve.

Figures 1–3 compare the true nonparametric regression function $g_0(\cdot)$ and its nonparametric estimator for the case of $H(v) = 0$ when the sample sizes are 200, 700 and 1200, respectively. Figures 4–6 compare the true nonparametric regression function $g_0(\cdot)$ with its nonparametric estimator for the case of $H(v) = v$ when the sample sizes are 200, 700 and 1200, respectively. The solid line is $g_0(\cdot)$ and the dashed line is the nonparametric estimator. We cannot forecast the trace of the random walk $\{V_t\}$ because of its non-stationarity. Hence, we estimate the true regression function $g_0(\cdot)$ according to the scope of $\{V_t\}$ and we cannot estimate $g_0(\cdot)$ in other points out of the scope since there is not enough sample in the neighborhood of each of such points. That is why the scopes of the abscissa axis are different in Figures 1–6. We can also find that the performance of the nonparametric estimate of $g_0(\cdot)$ improves as the sample size increases.

5. An empirical application

We use monthly observations on the U.S. share price indices, long-term government bond yields and treasury bill rates from Jan/1957–Dec/2009. The data are obtained from the International Monetary Fund’s (IMF) International Financial Statistics (IFS). The share price series used is IFS Series 11162ZF. The long-term government bond yield, which is the 10-year yield, is from the IFS Series 11161ZF. The treasury bill rate is from IFS Series 11160CZF. Figure 7(a)–(c) gives the data plots of the share prices, the long-term bond yields and the treasury bill rates.

To see whether there exist some statistical evidences for the three series to have the unit root type of non-stationarity, we carry out a Dickey–Fuller (DF) unit root test on

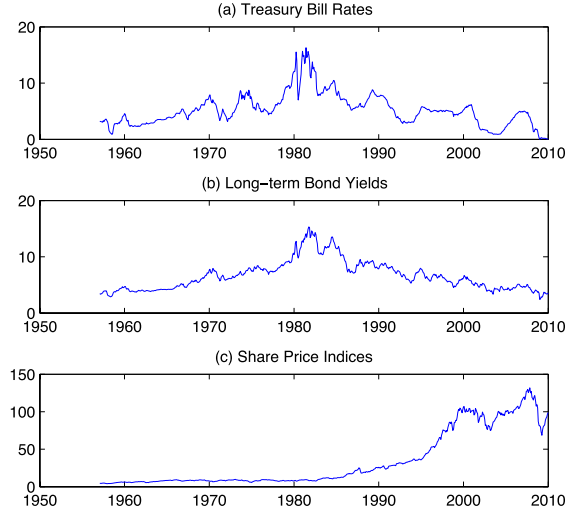


Figure 7. Time plots of the three series used in Section 5 over the period of Jan/1957–Dec/2009 with 624 observations. (a) treasury bill rates; (b) long-term bond yields; (c) share prices.

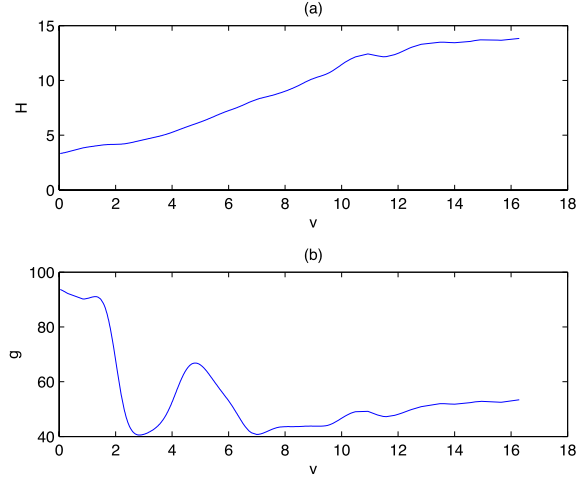


Figure 8. Estimates of the nonparametric functions $H(\cdot)$ and $g(\cdot)$ in Case A.

the three series. We first fit the data by an AR(1) model of the form

$$Z_t = \rho Z_{t-1} + e_t,$$

where Z_t = share price at time t or long-term bond yield at time t or treasury bill rate at time t . Then, by using the least-squares estimation method, we estimate the parameter ρ for the three series: for the share price series, $\hat{\rho}_{\text{share}} = 1.0023$; for the long-term bond yield series, $\hat{\rho}_{\text{Lbond}} = 0.9992$; and for the treasury bill rate series, $\hat{\rho}_{\text{Tbill}} = 0.9966$. Then we calculate the Dickey–Fuller t statistics and compare them with the critical values at the 5% significance level. The simulated P values for the long-term bond yields, treasury bill rates and share prices are 0.7040, 0.3130 and 0.4410, respectively. In addition, we also employ an augmented DF test and the nonparametric test proposed in Gao *et al.* [11] for checking the unit root structure of $\{Z_t\}$. The resulting P values are very similar to those obtained above.

Therefore, both the estimation results and the simulated P values suggest that there is some strong evidence for accepting the null hypothesis that a unit root structure exists in these series at the 5% significance level.

We then consider the following modelling problem:

$$\begin{aligned} Y_t &= X_t \theta_0 + g(V_t) + \epsilon_t, \\ X_t &= H(V_t) + U_t, \end{aligned}$$

where Case A: Y_t is the share price, X_t is the long-term bond yield and V_t is the treasury bill; and Case B: Y_t is the long-term bond yield, X_t is the share price and V_t is the treasury bill.

For Case A, the resulting estimator of θ_0 is $\hat{\theta} = -3.2155$ and the plots of the estimates of $g(\cdot)$ and $H(\cdot)$ are given in Figure 8. For Case B, the resulting estimator of

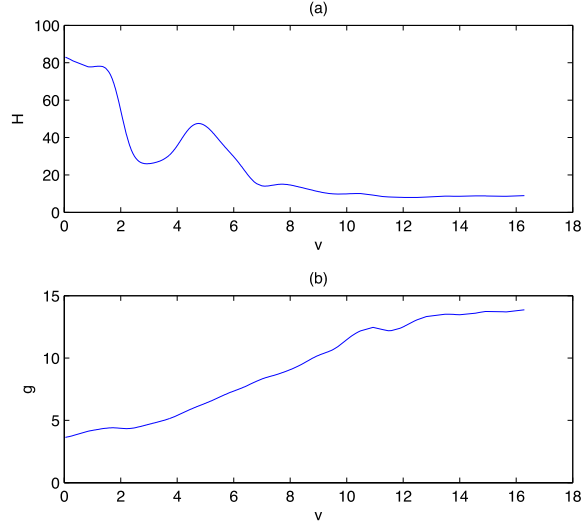


Figure 9. Estimates of the nonparametric functions $H(\cdot)$ and $g(\cdot)$ in Case B.

θ_0 is $\hat{\theta} = -0.0037$ and the plots of the estimates of $g(\cdot)$ and $H(\cdot)$ are given in Figure 9.

Figures 8 and 9 show that increases in treasury bill rates tend to lead to increases in long-term bond yields and decreases in share prices. Such findings are supported by the theory of finance and consistent with existing studies. Moreover, Figures 7–9 clearly indicate our new findings that both null recurrent non-stationarity and nonlinearity can be simultaneously exhibited in the share price, the long-term bond yield and the treasury bill rate variables.

Due to the cointegrating relationship among the stock price, the treasury bill rate and the long-term bond yield variables, our experience suggests that models (3.4) and (3.5) might be more suitable for this empirical study. We will have another look at this data after models (3.4) and (3.5) have been fully studied.

6. An outline of the proofs of the theorems

In this section, we provide only one key lemma and then an outline of the proofs of Theorems 3.1 and 3.2. The detailed proofs of the theorems are available from the supplemental document by Chen, Gao and Li [5].

Lemma 6.1. *Under the conditions of Theorem 3.1, we have as $n \rightarrow \infty$,*

$$\frac{1}{n} \tilde{X}^T \tilde{X} \xrightarrow{P} \Sigma. \quad (6.1)$$

Proof of Theorem 3.1. In view of Lemma 6.1 and the decomposition

$$\begin{aligned}\tilde{X}^\tau \tilde{X}(\hat{\theta}_n - \theta_0) &= \tilde{X}^\tau(\tilde{Y} - \tilde{X}\theta_0) \\ &= \sum_{t=1}^n \tilde{X}_t \tilde{g}(V_t) F_t + \sum_{t=1}^n \tilde{X}_t \epsilon_t F_t - \sum_{t=1}^n \tilde{X}_t F_t \left(\sum_{k=1}^n w_{nk}(V_t) \epsilon_k \right),\end{aligned}$$

in order to prove Theorem 3.1, we need only to show that for large enough n

$$\sum_{t=1}^n \tilde{X}_t \tilde{g}(V_t) F_t = o_P(\sqrt{n}), \quad (6.2)$$

$$\sum_{t=1}^n \tilde{X}_t F_t \left\{ \sum_{k=1}^n w_{nk}(V_t) \epsilon_k \right\} = o_P(\sqrt{n}), \quad (6.3)$$

$$n^{-1/2} \sum_{t=1}^n \tilde{X}_t \epsilon_t F_t \xrightarrow{d} N(0, \Sigma_{\epsilon, U}), \quad (6.4)$$

where $\tilde{g}(V_t) = g(V_t) - \sum_{k=1}^n w_{nk}(V_t)g(V_k)$. Recall that $\tilde{X}_t = X_t - \sum_{s=1}^n w_{ns}(V_t)X_s = U_t - \sum_{s=1}^n w_{ns}(V_t)U_s + \tilde{H}(V_t)$, where $\tilde{H}(V_t) = H(V_t) - \sum_{s=1}^n w_{ns}(V_t)H(V_s)$.

In order to prove (6.2)–(6.4), it suffices to show that for large enough n

$$\sum_{t=1}^n U_t \tilde{g}(V_t) F_t = o_P(\sqrt{n}), \quad (6.5)$$

$$\sum_{t=1}^n \bar{U}_t \tilde{g}(V_t) F_t = o_P(\sqrt{n}), \quad (6.6)$$

$$\sum_{t=1}^n \tilde{g}(V_t) \tilde{H}(V_t) F_t = o_P(\sqrt{n}), \quad (6.7)$$

$$\sum_{t=1}^n U_t \bar{\epsilon}_t F_t = o_P(\sqrt{n}), \quad (6.8)$$

$$\sum_{t=1}^n \bar{U}_t \bar{\epsilon}_t F_t = o_P(\sqrt{n}), \quad (6.9)$$

$$\sum_{t=1}^n \tilde{H}(V_t) \bar{\epsilon}_t F_t = o_P(\sqrt{n}), \quad (6.10)$$

$$\sum_{t=1}^n \bar{U}_t \epsilon_t F_t = o_P(\sqrt{n}), \quad (6.11)$$

$$\sum_{t=1}^n \tilde{H}(V_t) \epsilon_t F_t = o_P(\sqrt{n}), \quad (6.12)$$

$$n^{-1/2} \sum_{t=1}^n U_t \epsilon_t F_t \xrightarrow{d} N(0, \Sigma_{\epsilon, U}), \quad (6.13)$$

where $\bar{U}_t = \sum_{s=1}^n w_{ns}(V_t) U_s$ and $\bar{\epsilon}_t = \sum_{s=1}^n w_{ns}(V_t) \epsilon_s$.

In the following, we verify equations (6.5)–(6.13) to complete the proofs of Theorem 3.1(i) and Theorem 3.1(ii). Note that, for Theorem 3.1(i), equations (6.7), (6.10) and (6.12) hold trivially.

By the continuity of $g(\cdot)$ and $g'(\cdot)$, we have for $1 \leq t \leq n$,

$$\begin{aligned} & \frac{1}{N(n)h} \sum_{j=1}^n K\left(\frac{V_j - V_t}{h}\right) (g(V_j) - g(V_t)) \\ &= \frac{g'(V_t)}{N(n)h} \sum_{j=1}^n (V_j - V_t) K\left(\frac{V_j - V_t}{h}\right) (1 + o_P(1)). \end{aligned} \quad (6.14)$$

Thus, in view of (6.14) and Lemma 3.4 of Karlsen and Tjøstheim [20], in order to prove (6.5), it suffices to show that for n large enough

$$\sum_{t=1}^n U_t \Delta_n(V_t) F_t = o_P(\sqrt{n}), \quad (6.15)$$

where $\Delta_n(V_t) = \frac{g'(V_t)}{n^{\beta-\eta} h p_n(V_t)} \sum_{j=1}^n (V_j - V_t) K\left(\frac{V_j - V_t}{h}\right)$.

This kind of procedure of replacing $N(n)$ by $n^{\beta-\eta}$ and ignoring a small-order term as involved in (6.14) will be used repeatedly throughout the proofs in Appendices B and C of the supplemental document.

We then may show that (6.6) holds. Similarly to (6.14) and (6.15), we need only to show that

$$\sum_{t=1}^n \hat{U}_t \Delta_n(V_t) F_t = o_P(\sqrt{n}), \quad (6.16)$$

where $\hat{U}_t = \frac{1}{n^{\beta-\eta} h p_n(V_t)} (\sum_{k=1}^n K(\frac{V_k - V_t}{h}) U_k)$.

The detailed derivations for (6.15) and (6.16) are available from Appendix B of the supplemental document. The detailed proofs of (6.8), (6.9), (6.11) and (6.13) are also available from Appendix B. This will complete the proof of Theorem 3.1(i).

We then may prove Theorem 3.1(ii) by completing the proofs of (6.7), (6.10) and (6.12), which are again available from Appendix B of the supplemental document. \square

Proof of Theorem 3.2. By the definition of $\hat{g}_n(v)$, we have

$$\begin{aligned} \hat{g}_n(v) - g(v) &= \sum_{t=1}^n w_{nt}(v) (Y_t - X_t \hat{\theta}_n) - g(v) \\ &= \sum_{t=1}^n w_{nt}(v) (\epsilon_t + g(V_t) - g(v)) + \sum_{t=1}^n w_{nt}(v) X_t (\theta_0 - \hat{\theta}_n). \end{aligned} \quad (6.17)$$

Let $\Phi_{n,1} = \sum_{t=1}^n w_{nt}(v)(\epsilon_t + g(V_t) - g(v))$ and $\Phi_{n,2} = \sum_{t=1}^n w_{nt}(v)X_t(\theta_0 - \hat{\theta}_n)$. Then, we have

$$\hat{g}_n(v) - g(v) = \sum_{t=1}^n w_{nt}(v)(Y_t - X_t\hat{\theta}_n) - g(v) = \Phi_{n,1} + \Phi_{n,2}. \quad (6.18)$$

Since $\{\epsilon_t\}$ is assumed to be stationary and α mixing, by Corollary 5.1 of Hall and Heyde [14] and an existing technique to deal with the bias term (see, e.g., the proof of Theorem 3.5 of Karlsen *et al.* [19]), we have as $n \rightarrow \infty$

$$\sqrt{\sum_{t=1}^n K\left(\frac{V_t - v}{h}\right)} \Phi_{n,1} \xrightarrow{d} N\left(0, \sigma^2 \int K^2(u) du\right). \quad (6.19)$$

By (6.17)–(6.19), it is sufficient to show that

$$\sqrt{\sum_{t=1}^n K\left(\frac{V_t - v}{h}\right)} \Phi_{n,2} = o_P(1). \quad (6.20)$$

The proof of (6.20) may then be completed by Theorem 3.1 and Assumptions A1–A6. The details are available from Appendix B of the supplemental document. This completes an outline of the proofs of Theorems 3.1 and 3.2. \square

Acknowledgements

This work was started when the first and third authors were visiting the second author in 2006/2007. The authors would all like to thank the Editor, the Associate Editor and two references for their constructive comments on an earlier version. Thanks also go to the Australian Research Council Discovery Grants Program for its financial support under Grant Numbers DP0558602 and DP0879088.

Supplementary Material

Proofs of the theorems (DOI: [10.3150/10-BEJ344SUPP](https://doi.org/10.3150/10-BEJ344SUPP); .pdf). We provide this supplemental document in case the reader may want to have a look at the detailed proofs of Theorems 3.1 and 3.2 and Lemma 6.1. The details are available from Chen, Gao and Li [5].

References

- [1] BANDI, F.M. and PHILLIPS, P.C.B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** 241–283. [MR1956859](#)

- [2] BHATTACHARYA, P.K. and ZHAO, P.L. (1997). Semiparametric inference in a partial linear model. *Ann. Statist.* **25** 244–262. [MR1429924](#)
- [3] BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y. and WELLNER, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins Univ. Press. [MR1245941](#)
- [4] CHEN, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16** 136–146. [MR0924861](#)
- [5] CHEN, J., GAO, J. and LI, D. (2010). Supplement to “Estimation in semi-parametric regression with non-stationary regressors.” DOI: [10.3150/10-BEJ344SUPP](#).
- [6] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics VI* (J.J. HECKMAN and E.E. LEAMER, eds.) 5549–5632. Amsterdam: Elsevier.
- [7] ENGLE, R., GRANGER, C., RICE, J. and WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.
- [8] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. London: Chapman and Hall. [MR1383587](#)
- [9] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods*. New York: Springer. [MR1964455](#)
- [10] GAO, J. (2007). *Nonlinear Time Series. Semiparametric and Nonparametric Methods*. *Monographs on Statistics and Applied Probability* **108**. Boca Raton, FL: Chapman and Hall/CRC. [MR2297190](#)
- [11] GAO, J., KING, M., LU, Z. and TJØSTHEIM, D. (2009). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* **25** 1869–1892. [MR2557585](#)
- [12] GAO, J., KING, M.L., LU, Z. and TJØSTHEIM, D. (2009). Specification testing in nonstationary time series autoregression. *Ann. Statist.* **37** 3893–3928.
- [13] GREEN, P.J. and SILVERMAN, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. *Monographs on Statistics and Applied Probability* **58**. London: Chapman & Hall. [MR1270012](#)
- [14] HALL, P. and HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press. [MR0624435](#)
- [15] HÄRDLE, W., LIANG, H. and GAO, J. (2000). *Partially Linear Models*. Heidelberg: Physica-Verlag. [MR1787637](#)
- [16] HÄRDLE, W., LÜTKEPOHL, H. and CHEN, R. (1997). A review of nonparametric time series analysis. *Internat. Statist. Rev.* **65** 49–72.
- [17] HASTIE, T.J. and TIBSHIRANI, R.J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. London: Chapman and Hall. [MR1082147](#)
- [18] JUHL, T. and XIAO, Z. (2005). Partially linear models with unit roots. *Econometric Theory* **21** 877–906. [MR2168183](#)
- [19] KARLSEN, H.A., MYKLEBUST, T. and TJØSTHEIM, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.* **35** 252–299. [MR2332276](#)
- [20] KARLSEN, H.A. and TJØSTHEIM, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29** 372–416. [MR1863963](#)
- [21] LI, Q. and RACINE, J.S. (2007). *Nonparametric Econometrics. Theory and Practice*. Princeton, NJ: Princeton Univ. Press. [MR2283034](#)
- [22] LIN, Z. and LU, C. (1996). *Limit Theory for Mixing Dependent Random Variables*. *Mathematics and Its Applications* **378**. Dordrecht: Kluwer Academic. [MR1486580](#)

- [23] LINTON, O. (1995). Second order approximation in the partially linear regression model. *Econometrica* **63** 1079–1112. [MR1348514](#)
- [24] NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. *Cambridge Tracts in Mathematics* **83**. Cambridge: Cambridge Univ. Press. [MR0776608](#)
- [25] PHILLIPS, P.C.B. and PARK, J. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper No. 1181, Yale Univ.
- [26] ROBINSON, P.M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207. [MR0732897](#)
- [27] ROBINSON, P.M. (1988). Root- N -consistent semiparametric regression. *Econometrica* **56** 931–954. [MR0951762](#)
- [28] ROBINSON, P.M. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Rev. Econom. Stud.* **56** 511–534. [MR1023837](#)
- [29] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. U.S.A.* **42** 43–47. [MR0074711](#)
- [30] SCHIENLE, M. (2008). Nonparametric nonstationary regression. Working paper, Department of Economics, Univ. Mannheim, Germany.
- [31] SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall. [MR0848134](#)
- [32] WANG, Q. and PHILLIPS, P.C.B. (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25** 710–738. [MR2507529](#)
- [33] WANG, Q. and PHILLIPS, P.C.B. (2009). Structural nonparametric cointegrating regression. *Econometrica* **77** 1901–1948. [MR2573873](#)
- [34] WITHERS, C.S. (1981). Conditions for linear processes to be strong-mixing. *Z. Wahrsch. Verw. Gebiete* **57** 477–480. [MR0631371](#)

Received May 2010 and revised October 2010